# FUNCTIONAL EQUATIONS IN THE THEORY OF DYNAMIC PROGRAMMING—XI: LIMIT THEOREMS

Richard Bellman
Mathematics Division
The RAND Corporation

P-1843

30 November 1959



Reproduced by

The RAND Corporation • Santa Monica • California

## SUMMARY

In this paper we wish to present a limit theorem valid for a general class of Markovian decision processes. The result is of interest because of the simple conditions which are imposed and the rather simple argument which is used.

# FUNCTIONAL EQUATIONS IN THE THEORY OF DYNAMIC PROGRAMMING--XI: LIMIT THEOREMS

Richard Bellman

## 1. Introduction

In this paper we wish to present a limit theorem valid for a general class of Markovian decision processes, 1 . The result is of interest because of the simple conditions which are imposed and the rather simple argument which is used.

Let $p$ be an element of a finite set $P$, and $q$ be an element of another finite set $Q$. We think of $p$ as the state vector of a discrete dynamic programming process, and $q$ as the decision variable at each stage. A choice of $q$ results in a transformation from $p$ to $T(p,q)$, taken to be an element of $P$, and in a return of $b(p,q)$, a scalar function defined for all $p$ $P$ and $q$ $Q$.

Denoting by $p_1, p_2, \ldots, p_N$ the succession of states, and by $q_1, q_2, \ldots, q_N$ the sequence of decisions, we have as the overall return of an N-stage process the function

$$(1) \qquad R_N = b(p_1,q_1) + b(p_2,q_2) + \cdots + b(p_N,q_N).$$

We wish to choose the $q_i$ so as to maximize $R_N$.

Introducing the function $f_N(p_1)$ defined by the relation

$$(2) \qquad f_N(p_1) = \operatorname*{Max}_{q} R_N$$

for all $p_1$ $P$ and $N = 1,2,\ldots,$ we have the recurrence relation

$$(3) \qquad f_N(p_1) = \operatorname*{Max}_{q_1} \; b(p_1,q_1) + f_{N-1}(T(p_1,q_1)) ,$$

for $N \geq 2$, with

$$(4) \qquad f_1(p_1) = \operatorname*{Max}_{q_1} b(p_1,q_1).$$

It is reasonable to expect a "steady-state" policy which is approached asymptotically as $N \to \infty$; cf. [2,3,4], for results of this nature. The study of the asymptotic behavior of the sequence $f_N(p_1)$ determined by (3) is a problem of some difficulty, and usually requires some detailed knowledge of the transformation $T(p,q)$ and the function $b(p,q)$. We shall show in what follows that a fairly general result can be easily obtained under mild assumptions. Unfortunately, although we can derive the asymptotic form of $f_N(p)$, we cannot assert the existence of an asymptotic policy. Further assumptions appear to be required for this.

## 2. Statement of Result

Let us make the following two assumptions:

(1) (a) $b(p,q) \geq 0$, $p \quad P$, $q \quad Q$,

(b) $T(p,q)$ is such that by means of a suitable choice of $q$'s, $q_1, q_2, \ldots, q_K$, it is possible to go from any element $p_1 \quad P$ to any other element $p_2 \quad P$.

We wish to establish

<u>Theorem. Under the foregoing assumptions, for all</u> $p_1 \quad P$,

(2) $$f_N(p_1) \sim Na,$$

as $N \to \infty$, where $a$ is independent of $p_1$.

## 3. Proof of Theorem

Referring to (1.1), we may write

( ) $$f_{m+n}(p_1) = \max_{[q_1,q_2,\ldots,q_m]} \Big[ b(p_1,q_1) + \cdots + b(p_m,q_m) + f_n(T_m) \Big],$$

where $T_m$ is the state attained after the choice of $q_1,q_2,\ldots,q_m$.

Introduce the new sequence $\{u_n\}$ by means of the relation

(2) $$u_n = \max_p f_n(p).$$

Then, it is clear from (1) that

(3) $$u_{m+n} \leq u_m + u_n$$

for $m,n \geq 1$. It is well known that this inequality implies that there exists a constant $a$ such that

(4) $$u_n \sim na$$

as $n \to \infty$, [5].*

Let us now show that $f_n(p_1) \sim na$ as $n \to \infty$. Let for each $n$, $p_n$ be a value of $p$ for which $f_n(p)$ assumes the value $\max_p f_n(p)$. Choose a sequence of $q$'s, $q_1,q_2,\ldots,q_K$,

---

*This result is used in the foregoing fashion by Furstenburg and Kesten in a forthcoming paper.

which transforms $p_1$ into the value $p_{n-M}$. We know, by assumptions, that the number of transformations required to go from any point $p_1$ to any other point is uniformly bounded. Take $M$ to be this bound.

By virtue of the nonnegativity of $b(p,q)$, we have

(5) $$f_n(p_1) \geq f_{n-K}(p_{n-M}) \geq f_{n-M}(p_{n-M}).$$

Since $f_n(p_1) \leq f_n(p_n)$, by definition of the element $p_n$, we have for large $n \geq n( \ )$,

(6) $$n(a + \ ) \geq f_n(p_n) \geq f_n(p_1) \geq f_{n-M}(p_{n-M}) \geq (n - M)(a - \ ).$$

Hence

(7) $$f_n(p_1) \quad na$$

as $n \to \infty$, the desired result.

## REFERENCES

1. R. Bellman, Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1957.

2. R. Bellman, "On a quasi-linear equation," Canadian J. Math., vol. 8, 1956, pp. 198-202.

3. R. Bellman, "A Markovian decision process," J. Math. and Mech., vol. 6, 1957, pp. 679-684.

4. R. Howard, Discrete Dynamic Programming, John Wiley and Sons, New York, 1960.

5. G. Polya and G. Szego, Aufgaben und Lehrsatze..., Dover Publications, New York, vol. 1, p. 17, Exercise 98, 1945.